Nebraska Rural Independent Companies' Capital Expenditure Study

Predicting the Cost of Fiber to the Premise

January 2011

Contents

I. The Task

Under sponsorship from the Nebraska Rural Independent Companies,[1] several firms were engaged to produce a statistical procedure that could predict outside plant capital expenditures of a high-capacity terrestrial wireline broadband network.

The primary resource was a set of cost estimates produced by Vantage Point Solutions, Inc. (VPS), an engineering company located in Mitchell, South Dakota. These data reflected the engineering estimates of Fiber-To-The-Premises (FTTP) construction by incumbent local exchange carriers. The cost data were analyzed and compared to a variety of objective publicly available geographic and demographic variables (public variables). The goal was to produce a statistically reliable cost predictor with inputs limited to these public variables.

Analytical work during the project was performed by VPS,[2] by Consortia Consulting of Lincoln, Nebraska,[3] and by Rolka, Loube, Saltzer Associates.[4] GIS work was performed by Stone Environmental of Montpelier, Vermont.[5]

II. The Cost Data

The analysis began with the VPS estimates. Each case estimated the costs of constructing FTTP facilities within an entire exchange area.[6] The exchange areas generally consisted of small towns and the surrounding rural areas.

---

[1] The Nebraska Rural Independent Companies are: Arlington Telephone Company, Blair Telephone Company, Cambridge Telephone Co., Clarks Telecommunications Co., Consolidated Telephone Company, Consolidated Telco, Inc., Consolidated Telecom, Inc., The Curtis Telephone Company, Eastern Nebraska Telephone Company, Great Plains Communications, Inc., Hamilton Telephone Company, Hartington Telecommunications Co., Inc., Hershey Cooperative Telephone Company, Inc., K & M Telephone Company, Inc., The Nebraska Central Telephone Company, Northeast Nebraska Telephone Company, Rock County Telephone Company, Stanton Telephone Co., Inc., and Three River Telco.

[2] Most of the work for VPS was done by Larry Thompson, CEO, Quentin Flippin, PE, and Brian Enga, PE.

[3] Edit Kranner consulted for Consortia.

[4] Peter Bluhm and Dr. Robert Loube consulted for RLSA.

[5] David Healy and Katie Budreski consulted for Stone.

[6] Engineering projects that produced estimated costs for partial exchanges were excluded from the study.

A. Costs Covered

All locations within the relevant geographic area were served. A "location" could be a home, business or multiple dwelling unit. VPS recognized that each location could have more than one subscriber although in these regions a single location usually corresponds to a single subscriber.

The cost estimates covered both electronics and cable and wire outside plant. Generally, the estimates assumed reuse of all existing assets except the local loop.[7]

- Mainline fiber facilities were designed to each customer's drop point.
    - New fiber runs were included to remote platforms that were not already served by fiber.
    - Networks were designed to comply with current FTTP distance limitations (i.e. loop lengths were designed to be less than 12.4 miles.)
    - Mainline cables were often sized to serve empty lots because the clients generally had a duty to serve future growth.[8]
    - Most fiber was constructed using direct burial, although conduits were used for some in-town areas. The use of aerial plant for construction was insignificant.
- Drops were included if the customer currently subscribed to the client ILEC's service.
- Most cases involved replacing buried plant; therefore, the costs of removing existing copper feeder and distribution plant was not included because rarely is the cable removed.
- No cost was included for central office building modifications, although cost estimates included additional equipment for these central offices, such as fiber frames and FTTP Optical Line Terminals.
- Middle mile costs and inter-office costs were not included.

---

[7] From the point of view of feeder and distribution plant, the cost projections are essentially "greenfield" construction costs.

[8] Normally, VPS sized the mainline cable to accommodate any lot that had been platted and where there was a reasonable expectation that a house or business could be built on that lot. The incremental cost to add additional fibers (and possibly an access point) is small in comparison to the cost of new construction to add more mainline fiber when someone builds on that lot. On the other hand, fiber sizes generally did not provide for rural lots that had been abandoned, since these lots were considered unlikely to be occupied.

VPS cost estimates reflect current engineering standards and current cost factors at the time the work was done, which was in the period from 2004 to 2010.

In most exchanges, VPS prepared separate cost estimates for two portions of the exchange, one for the "town" area and the other for "rural" area. Costs were divided in this way because the cost for constructing a given distance of cable varies greatly between relatively congested towns and wide-open rural area. Town boundaries were sometimes defined by clients, but were often obvious. The "Town" classification was applied to areas with medium or high population density and a fine mesh road network. In all town cases, the area under study consisted of the entire town area. The "Rural" classification was applied to areas that were either clearly devoted to agricultural uses or undeveloped and had more widely spaced roads. Most exchanges produced a pair of records, one for the town area and one for the surrounding rural area.[9] We disregarded areas where separate cost data were not available for the town and rural areas.

With these limitations, the resulting database contained records for 227 rural areas and 209 town areas. These areas are in 15 states located primarily in the upper-central states and southeast, and they are served by 63 incumbent LECs. The data covered 54,000 route miles of mainline and drop route mileage and $1.103 billion of estimated construction costs.

Costs included labor, materials and engineering. For each cost record, VPS provided the following data:

- Central office FTTP electronics, optical network terminals (ONTs), spares, miscellaneous materials, and Central Office (CO) and ONT installation costs.
- Outside plant costs, including mainline optical cables, drops and fiber management equipment, with labor included.
- Engineering costs.
- Route miles for mainline cables and for drops.
- Number of locations served.
- Land area.

[9] The number of rural areas does not match the number of town areas because in some cases VPS was only directed to design facilities for the rural area.

B. Geo-referencing the Cost Data

The project's first GIS task was to associate each VPS cost record with a particular geographic area.[10] We derived the boundaries of these "cost areas" from several sources.

Telephone exchange boundaries were the most important data set. Although there are commercial sources for this information, we did not find them to be as current and reliable as other sources. In each state where VPS provided cost data, we relied on the best available sources for exchange boundaries.

- For Minnesota, Ohio, and Wisconsin we used data acquired from state agencies in GIS formats.
- For Colorado, Missouri, Montana, North Dakota, and Tennessee we downloaded exchange boundary maps from state telecommunications association web sites. We then geo-referenced and used on-screen digitizing to convert these maps to GIS formats.
- For Idaho and North Carolina, we obtained graphical maps from the state Public Utilities Commissions. We then geo-referenced and used on-screen digitizing to convert these maps to GIS formats.
- For Florida, Georgia, Iowa, Illinois, Indiana, Michigan, Nebraska, and South Dakota we compiled exchange boundaries from VPS CAD files.

Table 1 summarizes these sources and methods.

Table 1. Sources of Exchange Boundary Data by State

State	Data Type	Organization Source
CO	Digitized from Georeferenced Map by Stone	Colorado Telecommunications Association
FL	CAD Files	VPS
GA	CAD Files	VPS
IA	CAD Files	VPS
ID	Digitized from Georeferenced Map by Stone	Idaho Public Utilities Commission
IL	CAD Files	VPS
IN	CAD Files	VPS
MI	CAD Files (Exchange had to be stretched to area)	VPS
MN	State GIS Database	Minnesota State GIS
MO	Digitized from Georeferenced Map by Stone	Missouri Telecommunications Industry Association

[10] The GIS system represented the boundaries of each in-town area and each rural area as a polygon with specific geographic reference and a standard national projection system.

State	Data Type	Organization Source
MT	Digitized from Georeferenced Map by Stone	Montana Telecommunications Association
NC	Digitized from Georeferenced Map by Stone	North Carolina Utilities Commission
ND	CAD Files Digitized from Georeferenced Map by Stone	North Dakota Telephone Association
NE	CAD Files	VPS
OH	State GIS Database	Ohio State GIS
SD	CAD Files	VPS
TN	Digitized from Georeferenced Map by Stone	Tennessee Telecommunications Association
WI	State GIS Database	Wisconsin Public Service Commission

As described above, data for seven states were geo-referenced and digitizing at Stone. This process required several steps. First, we imported digital telephone exchange maps into the GIS. Then, we geo-referenced the maps to known geographic features such as roads and rivers. We also used other datasets to verify cost area boundaries, including administrative boundaries and information from VPS cost records. Finally, we used the geo-referenced digital maps to digitize the boundaries of the telephone exchange areas.

Because all of the VPS cost records had separate costs for "in-town" areas and outlying "rural" surrounding areas, within each telephone exchange it was necessary to establish geographic boundaries between the two. We used two methods.

- Some town areas have been defined by the Census as "populated places." For 142 telephone exchanges, we intersected the exchange centroid's latitude and longitude with the US Census Bureau's Populated Place GIS database. We then assigned a town area record's boundary to the known boundary for the Census Bureau's populated place. Finally, we checked the characteristics of the resulting areas against the VPS data on mainline mileage and number of locations.

- For the remaining 67 town exchanges, we used visual methods to establish town boundaries. Using the exchange centroid for location, we reviewed electronic photo-imagery from the 2007 National Agriculture Imagery Program (NAIP) to establish the boundaries of the settled areas. We then digitized those boundaries on-screen. Finally, we checked the characteristics of the resulting areas against the VPS data on mainline mileage and number of locations.

Having defined all the town and rural cost record boundaries digitally, we consolidated the information into a single GIS database.

III. Analogous Public Data

As stated above, the project goal was to produce a statistically reliable cost predictor that relies solely on public variables. First, we sought variables that might serve as proxies for three variables reported in the VPS cost data: area, locations served and mainline route miles. We discuss here the three GIS variables that we selected for this purpose. The remaining public data variables are discussed in part IV below.

Stone Environmental derived all of the project's public data from readily available government and public GIS data sources. To use these kinds of public variables required a process to select the appropriate data and then conform it to the geographic boundaries of the VPS cost records. As described in more detail below, for each variable Stone acquired the necessary public data and then developed a GIS Data Model for conducting the appropriate data extraction. In the end, each VPS cost record had assigned to it a single datum for each public variable.

A. Public Data Substitutes

1. GIS Area

The first public variable was area. Using GIS software, we calculated the GIS area from the compiled digital boundaries associated with each cost record.

2. Households

The second public variable was the number of households. The U.S. Census Bureau reports these data at the "census block" level. The census block is the smallest geographic unit for census data. In urban areas, a census block is roughly one city block. In rural areas, particularly sparsely populated areas, a census block can be a variety of sizes based on the number of housing units.[11] Our task for census data, therefore, was to define a procedure that would map the census block attributes to cost record boundaries.

We used the "centroid" method to estimate households from census data. Each census block has a geographic balance point known as its "centroid," with known coordinates. If the centroid of a census block fell within a cost area boundary, we attributed 100% of the households of that census block to the cost area. Conversely, if the centroid of the census block fell outside the cost area, we ignored the households within that census block.

While the centroid method is an efficient way to match the project's cost data with public variables, it is not exact. An exchange's attributes are overstated when an included census block has some area outside the cost area boundary. Conversely, an exchange's attributes are understated when an excluded census block has some area inside the cost area boundary. On average, the centroid method should not produce any systematic bias upward or downward, although it can introduce estimation errors for individual cases.

---

[11] In our set of cost records, the number of census blocks per record varied from one to 1,039. The average was 107.

3. Street Mileage

Our preliminary analysis of the VPS cost data showed that mainline route mileage per location served as an important cost driver. To find a proxy within public data, we obtained a GIS version of the nation's road network. We used a national database called "Streetmap" that comes with ESRI's ArcMap Software.[12] To define how many road miles exist within each cost area, we "clipped" road segments at cost record boundaries using the GIS program. Road segments and portions of road segments that lie outside exchange boundaries were excluded. This method is the most precise for estimating actual road mileage. As discussed in the next section, we later decided to adjust road mileage to account for unpopulated areas and for road types that are unlikely to support utility rights-of-way.

B. Qualifying and Adjusting the Data

As noted above, the variables derived from publicly available data included area, households, and road mileage. Each of these three variables had analogues in the VPS data. We used all three of these variables to validate the GIS exchange area boundaries against the VPS cost data boundaries and to exclude extreme cases where a geographic error or mismatch seemed likely.

1. The Area Gate

First, we calculated an area ratio equal to Exchange GIS area divided by VPS cost area (Area Ratio). If the Area Ratio was very high or very low, we produced maps of the area and visually verified that the perimeter of the exchange and the included census blocks were similar. This step allowed us to identify and correct some initial errors in cost area boundary data.

After reviewing the variation in this Area Ratio, we still found some cases with a large variance from the expected value of 1.00. We were concerned that this variation, because it was unexplained, might limit reliability in the subsequent regression analysis. We adopted a quality control gate that excluded cases for which the Area Ratio was less than 90% or more than 110%. This gate was intended to exclude all cases that have a difference in area arising from differences in exchange boundary placement or some other undiagnosed cause. Of the 436 records available, 391 passed through this first quality control gate.

2. The Household Gate

Second, we calculated the ratio of Census households to VPS locations (Location Ratio). The analysis of this ratio was more complex.

First, the VPS data included business locations, but the Census data for households do not. In commercial areas, therefore, one would expect the Location Ratio to be less than 1.0. In fact, the mean ratio for all cases was 0.87.

---

[12] "ESRI" is a company located in Redlands, California and is the publisher of a variety of GIS products. The original source of the Streetmap data was TeleAtlas, a private data reseller.

Second, there are timing differences. Location data from VPS were collected at the time of the engineering estimate, from 2004 through 2010. Census household data was derived from the 2000 census. In an area with an increasing population, one would expect the Location Ratio to be less than 1.0. Conversely, an area experiencing an outflow of population (as in many rural areas in the Midwest and Great Plains regions) would have a Location Ratio greater than 1.0.

We corrected for timing differences by developing an "adjusted households" datum for each record. We increased or decreased Census households reported for 2000 by the same percentage that population increased or decreased in the same county between 2000 and 2009. We tested both the unadjusted household and the adjusted household number for relevancy. As explained below, we ultimately adopted and used solely the unadjusted raw number of households.

We adopted a second quality control gate based on the unadjusted Location Ratio.[13] For the reasons explained above, the mean Location Ratio approximated 0.90. We therefore centered our gates on that value. We excluded cases if the Location Ratio was less than 0.70 or more than 1.10. This gate was intended to exclude cases that have a large error arising from the centroid approximation or from some other undiagnosed cause. Of the 436 records available, 297 passed through this second quality control gate.

3. The Road Mileage Gate

Third, we calculated the ratio of GIS road miles to mainline route miles recorded in the VPS cost data (Roads Ratio). This analysis was the most complex of the three because several factors prevent road distances from mapping closely to fiber route miles. Some roads exist in locations where a carrier would not build fiber. These include roads that serve seasonal dwellings and special purpose roads used for agriculture, fire protection, or other reasons. Also, if an area is laid out with a square grid of roads, it is often possible to serve all customers by building fiber routes along the north-south roads, with short spurs running out to locations on the east-west roads. Finally, certain road classes have historically not been used for utility rights of way. For all of these reasons, the GIS road miles were expected to be greater than the VPS mainline route miles, making the Roads Ratio greater than 1.0.

We made two adjustments to raw road distances. First, we corrected for unpopulated areas. To adjust for unpopulated areas, we identified census blocks with zero population, and we eliminated road segments that crossed those census blocks. This correction generally brought road miles closer to fiber route miles, but does not account for the occasional case where a carrier must build facilities across unpopulated areas to serve remote customers. The second adjustment was by road type. In the Streetmap database, each road segment has a defined road class. We summarized the road segment mileage by each class for each cost area.

- We excluded limited access federal highways, because it is not generally possible to build fiber along these roads.

---

[13] For reasons discussed below, we finally rejected the population adjustment of households because it did not improve results.

- We also excluded major divided highways, roads with special characteristics such as cul-de-sacs, access ramps, and traffic circles, and other thoroughfares such as walkways and driveways.[14]

These preliminary adjustments decreased the variance between the road mileage and route miles variables. Throughout the database, the average Roads Ratio with these adjustments was1.04.

We adopted a third quality control gate based on the Roads Ratio. Cases were excluded if this ratio was less than 0.80 or more than 1.20. This gate was intended to exclude cases that have a large error arising from the road clipping process or from some other undiagnosed cause. Of the 436 records available, 258 passed through this third quality control gate.

A cost record was excluded if it failed any of the three quality control gates. After applying all three gates, 168 records were used, comprising 86 rural cases and 82 town cases.

We made a final adjustment to the database by rejecting one outlier with a reported cost per location that was far above the regression line. When we checked against VPS records, we found a serious inconsistency. VPS records showed that this case had high costs due to rocky soils. The GIS soils data, however, did not identify unusual soil conditions in that area. We concluded that either the GIS data or the VPS data was in error, so we excluded the data point. The resulting database therefore contained 167 records.

IV. Other Cost Drivers

Stone Environmental collected a variety of GIS data from public sources to test in the regression study. The following paragraphs describe these public variables, how we processed that data using GIS technology, and how the final variable was defined for the regression analysis.

- Soils Texture. We postulated that areas with rocky soils and certain kinds of dense soils would have higher construction costs. We used the clipped and adjusted Streetmap data with Soil Survey Geographic Database (SSURGO) soils data from the Natural Resource Conservation Service (NRCS) for each cost record. We assigned the attributes of the most significant SSURGO soil polygon to each road segment within the cost area. Each soil was then rated for construction difficulty and assigned a construction cost multiple ranging from 1 to 3, with 1 representing the lowest cost soils and 3 representing the highest cost soils for telecommunications construction.[15] A mile of road with a rating of

---

[14] We excluded major divided highways (FCC Class 1), roads with special characteristics such as cul-de-sacs, access ramps, and traffic circles (FCC Class 6) and thoroughfares including walkways and driveways (FCC Class 7).

[15] The values for soil difficulty were the same as those used in the FCC's current Hybrid Cost Proxy Model.

3.0 therefore was assumed to cost three times as much to install fiber as a mile of road with a soil rating of 1.0. We then summed the weighted road results for each cost area and divided by the adjusted road length in that cost area. The resulting index reflects the average soil construction difficulty in the cost area.

- Bedrock Percentage. We postulated that areas with shallow bedrock would have higher construction costs. We intersected the clipped and adjusted Streetmap data with SSURGO soils data for each cost area. We assigned the attributes of the underlying SSURGO soil polygon to each road segment within the cost area. We identified road segments where the soils indicated a depth to bedrock of less than 36 inches, a common depth of plowing buried cable. We summarized for each cost record the road miles where depth to bedrock was less than 36 inches and divided by adjusted road length in that cost area. The resulting value is the percentage of roads in the cost area with shallow bedrock that would interfere with installation of fiber.

- Road Intersections Frequency. We postulated that road intersections slow fiber construction and impose other costs. We derived data from the clipped and adjusted ESRI Streetmap data using GIS technology. The regression input variable was the average number of road intersections within the cost area per mile of adjusted road length.

- Stream Crossings Frequency. We postulated that stream crossings slow construction and impose other costs. We used the USGS National Hydrography Dataset (NHDPlus) and the clipped and adjusted ESRI Streetmap data. The regression input variable was the number of stream crossings for roads within each cost area per mile of adjusted road length.

- Wetlands Percentage. We postulated that fiber construction in wetlands would require additional approvals and specialized techniques. Where available, we used US Fish and Wildlife Service National Wetland Inventory data to identify clipped and adjusted road segments that intersect with wetlands. For Colorado, Wisconsin, and Montana, those data were not available. For these states, we used the SSURGO soil drainage class as an indicator for the presence of wetlands. Where the drainage class was very poorly drained or poorly drained, road segments were classified as wetland. We then intersected the road miles with wetland areas and summarized by cost area. For each cost record, we summarized the road miles intersecting with wetland, and the result was divided by the adjusted road length. The regression input variable, therefore, was the average percentage of road miles within wetland areas.

- Frost Index. We postulated that costs would be higher in areas where the construction season is shorter due to ground freezing. We used the clipped and adjusted Streetmap data and applicable SSURGO soils data. The number of frost-free days for each underlying SSURGO soil polygon was then assigned to each road segment. The road segment values were then averaged over the cost area. The approximate mean of our data was 164 frost-free days per year. The regression input variable for any cost area, called here the "frost index," was 164 minus the number of frost-free days per year in that area, weighted by road mileage. Accordingly, a cold area will have a positive Frost Index, and a warm area would have negative Frost Index.[16]

- Rain Frequency. We postulated that areas with frequent rain would have more construction delays and higher costs. We identified the nearest Samson weather station for each cost record. We identified NOAA National Climate Data Center precipitation data from 1975 to 2000 for all Samson stations. The input statistic was the average number of days per year with greater than 0.5 inches of precipitation.

V. The Regression Study

After defining the input data and performing quality control checks on the GIS work, Rolka, Loube, Saltzer Associates and Consortia performed regression analyses using the public variables as independent variables and construction cost per household as the dependent variable.

A. Correcting Costs for Inflation

Because the cost studies were done over a period of years, all cost data were first updated to 2010 prices using the Consumer Price Index. The adjustment factors for cost were as follows:

Table 2. Inflation Adjusters

Construction Year[17]	CPI	Inflation Adjuster
2004	188.9	1.1578
2005	195.3	1.1199
2006	201.6	1.0849
2007	207.3	1.0548
2008	215.3	1.0158
2009	214.5	1.0195
2010	218.7	1.0000

B. Cost Relationships within the VPS Data

The VPS data showed several interesting features.

- For mainline cable in rural areas, the average total project construction cost (including all electronics, outside plant, and other costs as described on page 3) was $26,728 per mile. Average total project construction cost (including all

---

[16] For example, an area that has 174 days of frost per year would have a Frost Index of 10.

[17] The 2010 annual CPI had not been released at the time of publication. We used the October, 2010 CPI value.

electronics, outside plant, and other costs as described on page 3) in town areas was $192,931 per mile. VPS believes the in-town costs per mile were generally higher because town projects require more conduit, more frequent road crossings, more coordination with other utilities, and more frequent customer drops.

- Per-customer costs are quite different between rural and town areas, because of lower customer densities in rural areas. Average costs per customer were $4,438 in town areas and $9,286 in rural areas. This difference arises because rural customers require more mainline cable than town customers.

- Costs were unevenly distributed. The FCC has recently observed that a substantial portion of cost is incurred to serve a small minority of customers. This phenomenon held true for these construction projects, even though each project is itself an aggregate of customers with varying individual costs. The three most expensive cost areas, (representing 1.7% of the projects) consumed 12% of the total investment for the whole data set.

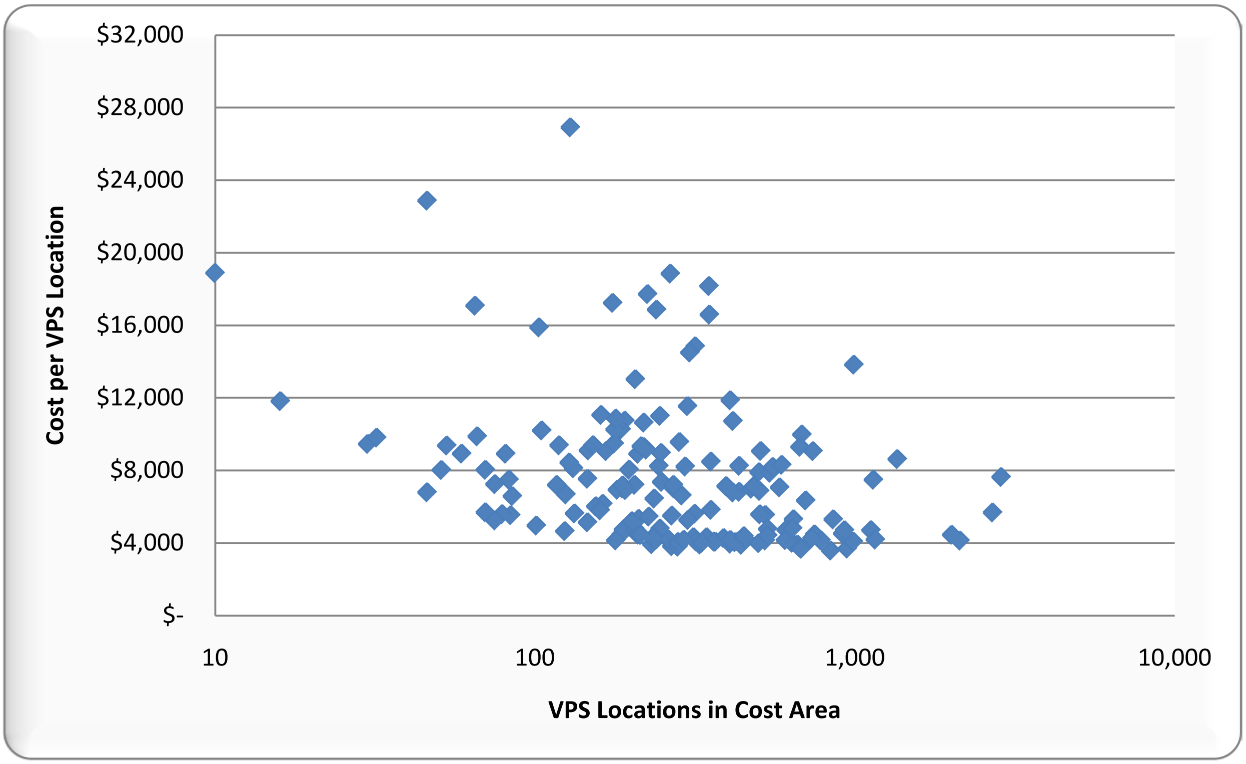
- Cable investment constitutes a significant portion of total plant investment. Outside plant comprised 58.5% percent of the total investment in the data set.

1. In-Town Cost per Location by Project Size

The VPS data showed generally that smaller projects had a higher minimum cost per location than larger projects.[18] The relationship is shown in Chart 1.

[18] In Charts 1 through 3, VPS data is limited to cases that passed through all three quality control gates.

Chart 1. Cost per Location by Size



Generally, as project size increased, the minimum cost converged to about $4,000 per customer, although there were many cases of higher cost. This preliminary result caused us to later examine project size as a potentially significant independent variable to predict cost.

2. Cost and Density

Our preliminary analysis of the VPS data showed that density is a strong predictor of cost. The relationship closely fits a hyperbola of the form:[19]

$$\frac{Cost}{Location} = a + \frac{b}{Density}$$

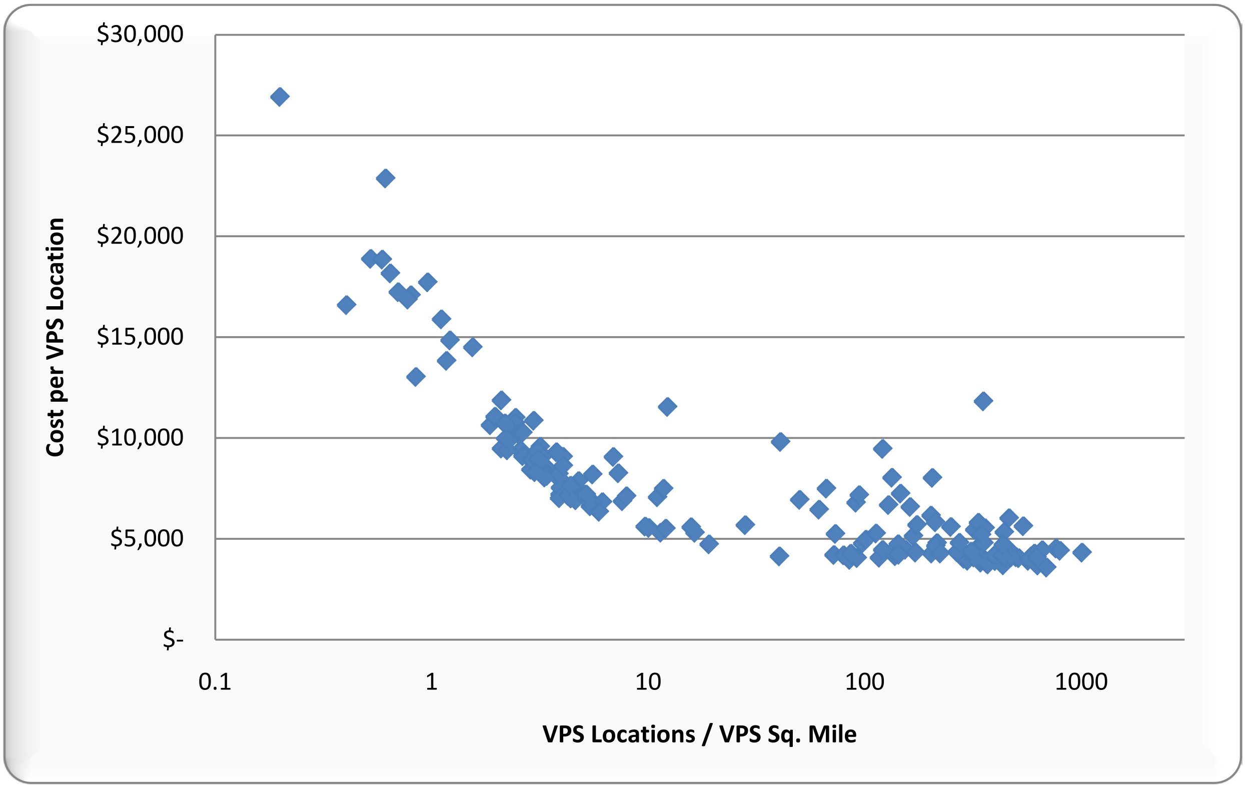
A threshold question was whether "density" in this formula is better defined on an area basis (customers/square mile) or a linear basis (customers/route mile). Traditional telecommunications cost models have used area density, yet area density tends to assume that area itself creates cost, an assumption that is untrue in unpopulated areas. Linear density also makes sense, particularly for wireline networks that have relatively uniform costs per route mile.

---

[19] To evaluate this kind of inverse relationship, we used 1/density as an input variable in our regression analysis.

The VPS data did show a strong association between cost per location and area density, as is shown in Chart 2:

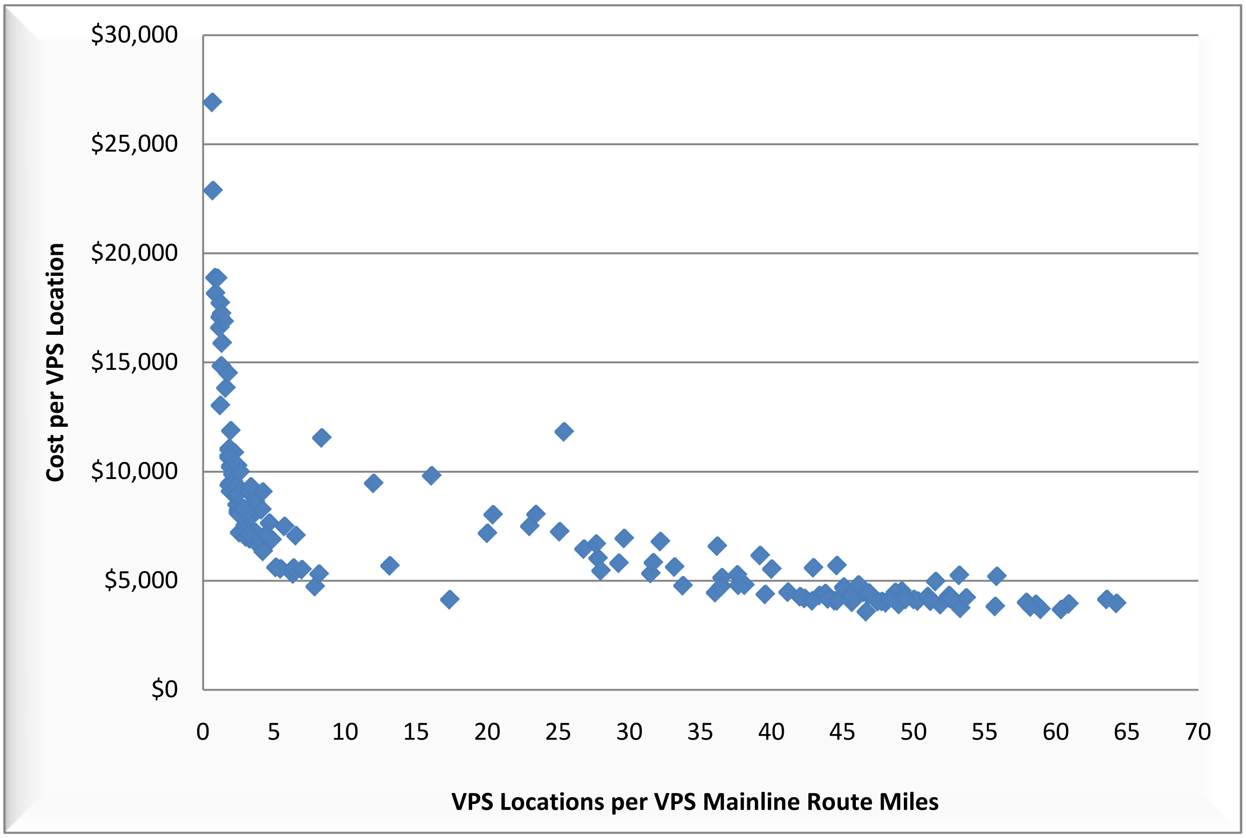
Chart 2. Cost per Location by Area Density



These data show a strong association, with some outlying cases. The relationship between cost per VPS location and VPS Location per VPS square mile has an R-squared of 0.71. In other words, area density can explain 71% of the variation in cost per location.

Chart 3 shows the pattern in the VPS data between cost per location and linear density.

Chart 3. Cost per Location by Route Density



In this case the R-squared is 0.87, meaning that linear density can explain 87% of the variation in cost per location. Because the R-squared is higher for linear density means, the relationship between linear density and cost per location is stronger, and all of the following regression analysis uses linear density.

The best-fitting curve for the VPS data has the following features:

$$\frac{Cost}{Location} = \$4{,}430 + \left[\frac{\$12{,}911}{Locations/_{Route\ Mile}}\right]$$

Algebraically, this simplifies to the following cost model that considers the two inputs variables recorded by VPS:

$$Cost = [\$4{,}430 * Locations] + [\$12{,}911 * Route\ Miles]$$

C. Substituting households and road miles for locations and route miles

Using the VPS locations and VPS route miles created two problems. First, if these data are to be used for USF distributions, new reporting requirements would have to be imposed requiring supported carriers to report locations served and mainline route miles. Second, even if such reporting requirements were created, they might generate poor incentives. For example, a high-cost support mechanism that increased support with route miles could create an incentive for inefficient network layouts. Therefore, we evaluated possible substitutes from public sources.

- First, we substituted GIS adjusted road miles for VPS mainline route miles. With this substitution, the R-squared dropped from 0.87 to 0.85. The difference was hardly noticeable, and we judged the substitution acceptable.

- Second, we substituted census households as a proxy for VPS locations. With this substitution, the R-squared dropped from 0.87 to 0.84. Census households were nearly as good at predicting cost as the VPS locations in the original data.

- Having tested the two substitutions separately, we tested both together. The resulting equation produced an R-squared of 0.825. This result is not substantially different from the R-squared of 0.87 from using only VPS data.

Therefore, the substitution of the public variables does not substantially degrade the ability of density to predict cost. We concluded that using Census households as a substitute for service locations and using adjusted road miles as a substitute for mainline route miles still allowed us to make a highly reliable cost prediction.

The household data was from the 2000 census. We fully tested both the raw household data and household data that had been adjusted to reflect trends in county populations from 2000 to 2009. Using adjusted households did not improve the reliability of the result. Therefore, either input could be used with equal validity. Based on the principle of parsimony, we decided to use the raw household data in all calculations involving households.

The cost equation at this point was:[20]

$$\frac{Cost}{Household} = \$5{,}042 + \left[\frac{\$13{,}134}{Households/_{Adjusted\ Road\ Miles}}\right]$$

Alternatively:

$$Cost = [\$5{,}042 * Households] + [\$13{,}134 * Adjusted\ Road\ Miles]$$

This striking result means that one can use two points of public source data (households and road mileage) to predict 82.5 percent of the variation in FTTP construction costs in town and rural areas.

[20] The results were highly significant. The T-test statistic for the density variable was 27.91.

D. Adding other public variables

As described above, we considered a variety of other variables as potential cost drivers. We also evaluated whether we could improve the accuracy of the cost model by considering these other geographic variables. Adding these variables did add incremental value to the reliability of the cost predictions.

The following table shows the effects of adding these other public variables, taken one at a time.

Table 3. Regression of Single Added Variables

Added Independent Variable	R-squared	Regression Equation Coefficient	T-Statistic
None	0.8252		
Households	0.8294	-0.8867	-2.02
Frost Index	0.8521	-32.76	-5.46
Wetlands Percentage	0.8419	30608	2.37
Soils Texture	0.8322	2559	2.62
Road Intersections Frequency	0.8398	252.6	3.87
Bedrock Percentage	0.8300	2769	2.17
Stream Crossings Frequency	0.8253	-1.37	-0.31
Rain Frequency	0.8283	-41.13	-1.73

When added individually to linear density, each of the first six variables listed in Table 3 was statistically significant: Households, Frost Index, Wetlands Percentage, Soils Texture, Road Intersections Frequency, and Bedrock Percentage. Our standard for statistical significance was that the extra variable increased the R-squared and had a T-statistic of greater than 2.0 or less than minus 2.0.

We then conducted a multivariate regression. In this analysis, our standard for significance required more judgment. We included a variable after considering whether its inclusion increased the R-squared of the regression, whether the sign of its coefficient was as we expected, and whether the T-statistic was 1.28 or larger.[21]

The same variables turned out to be significant as above, except soils and bedrock. The Soils Texture variable and the Bedrock Percentage variable do not purport to measure the same thing. The soils variable purports to measure the difficulty of plowing cable into soil, and bedrock measures the frequency that no soil exists at plow depth. Nevertheless, in this analysis the two variables behaved very similarly. Although each was significant when considered by itself, neither was significant when the two were deployed together in the multivariate analysis. This result was contrary to the expectation of our engineers, and may justify further study. It may be that a larger sample of cost data, covering a wider range of soil conditions and regions

[21] For a large number of observations, a T statistic of 1.28 is considered significant at the 90% confidence level.

of the country, could establish both of these variables as significant in a multivariate analysis. Another possibility is that our protocol for estimating soil difficulty should be revised.

We selected soils as the preferred variable, although the other decision would have been almost equally valid. Both raised the R-squared to approximately the same level, and both had T-statistics of approximately equal value.

The final regression equation takes the following form, with the coefficients defined in Table 4:

$$\frac{Cost}{Household} = [A] + \left[\frac{B}{Households/_{Adjusted\ Road\ Mile}}\right] + [C * Households] + [D * Frost\ Index] + [E * Wetlands\ Pct.] + [F * Soils\ Texture] + [G * Road\ Intersect.Freq.]$$

Table 4. Multi-Factor Regression Coefficients

Factor	Coefficient Symbol	Coefficient	T-statistic
Fixed cost	A	$3,072	Not applicable
Linear Density	B	$13,365	18.96
Households	C	-$0.8867	-2.10
Frost Index	D	$25.04	3.61
Wetlands Pct.	E	$17,700	1.38
Soils Texture	F	$1,376	1.49
Road Intersections Frequency	G	$165.40	2.46

The final R-squared of the seven-term regression was 0.8666. This means that if one knows for a particular area the number of households and road miles, as well as the frost, wetlands, soils and road intersections, one can explain 86% of the variation in the cost of constructing FTTP facilities.

VI. Significance for universal service and cost modeling

It cannot be disputed that rate-of-return regulation has been a broadband success story. On a per-capita basis, rate-of-return companies have more broadband deployment per capita than areas served by non-rural ILECs that are subject to price caps, even though rate-of-return companies undoubtedly have a greater portion of their customers located in very rural areas.

Nevertheless, it is understood that the FCC is looking for ways to manage the size of the high cost fund as it is transitioned into a broadband mechanism. The results of this analysis, although preliminary, could be useful to the FCC in determining methodologies for capital expenditures necessary to construct terrestrial-based broadband facilities. Among the potential uses of the results that may be considered:

- Developing a mathematically supported upper limit on "reasonable" capital expenditures based on linear density and potentially other factors that are determined to be significant.
- Developing a methodology for limiting the pace of broadband facilities deployment by companies in light of funding limitations.
- Evaluating the nationwide "cost" of deploying high-capacity terrestrial broadband networks.

Insofar as utilizing the results to determine caps, one possibility is to set caps on the amount of investment or the rate of investment for rate-of-return companies, given that the results explain to a very high percentage the variance in costs among FTTP installations. While the raw data reflected projects by a single engineering firm, those projects numbered in the hundreds and covered numerous states across a wide swath of geography. While not all carriers are installing FTTP for all customers, these data can be instructive in deciding a range of costs. Since the formula for these FTTP costs represents the mean value, however, another formula would need to be statistically determined to establish an upper boundary to the cost range.

The high power of these public variables – particularly linear density – to predict costs strongly suggests that it would be useful to conduct further analysis with a larger data set, possibly with engineering data from other firms and other states. While the data here were geographically diverse, seeking more cases from mountainous areas and coastal regions would make the results even more robust. Even as the study stands now, because it already has yielded such a reliable cost predictor of broadband construction, the study should be useful to the FCC in refining its nationwide cost estimates for broadband construction. In light of the scarcity of available high-cost support, it is tremendously important to achieving the nation's broadband deployment goals that determining broadband costs is done as accurately as possible.

Appendix I – Stone Environmental Inc. GIS Protocols

STONE ENVIRONMENTAL INC

535 Stone Cutters Way
Montpelier, Vermont
05602 USA

Phone / 802.229.4541
Fax / 802.229.5417
Web Site / www.stone-env.com

Project: 102413-G **Date:** 10/28/10, rev. 12/13/2010

Subject: Fiber to Home GIS Data Processing and Extraction

PURPOSE/OBJECTIVE:

Develop tools for extracting information related to expenditures of a high capacity terrestrial wireline broadband network.

PROCEDURE:

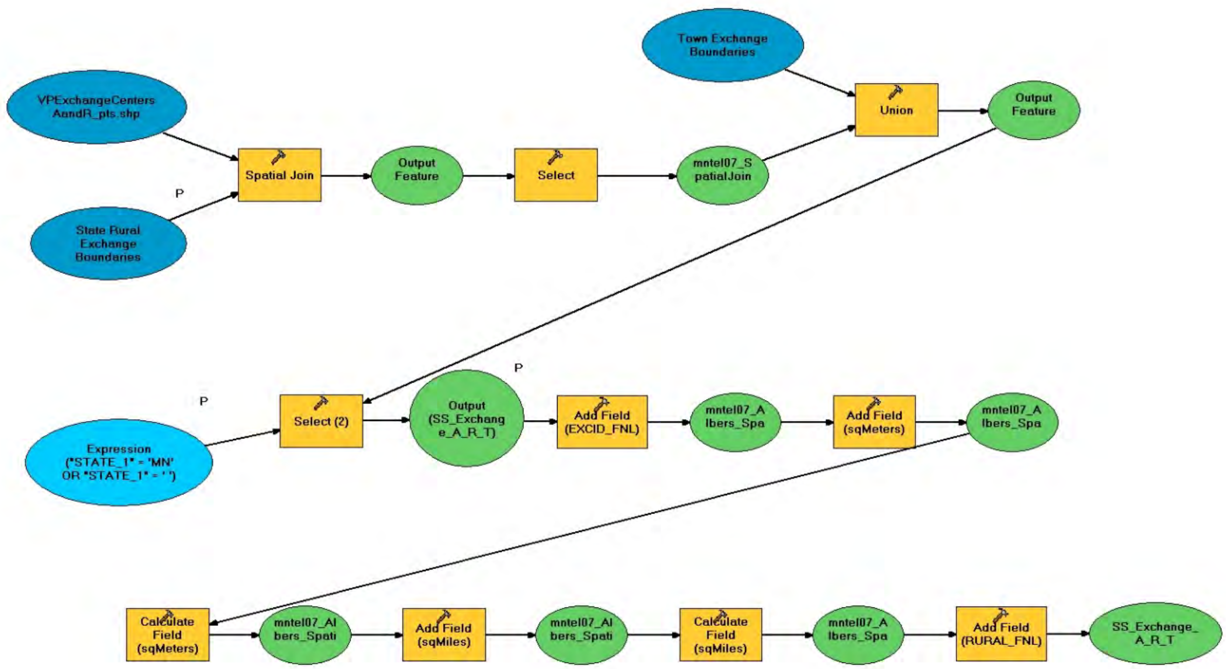
All Tools are in the following location: Tools **O:\Proj-10\AIM\2413-G-FibertoHome\Data\GISData\Tools\FiberToHome.tbx**

1) **Exchange Boundary Processing: Tool – 0_PrepareExchangesByState**
   a. *Tool description:* Intersects town based polygons with rural and all exchange boundaries and assigns a secret code to each of the Telephone Exchange Boundaries.
      i. Spatially joins rural exchange boundaries with exchange centers to assign exchange information to exchange boundaries (polygons).
      ii. Unions rural exchange boundaries for state of interest with town exchange boundaries.

iii. Add fields to store exchange information: ExcID, sqMeters, sqMiles

iv. Calculate area of exchanges:

1. Square Meters (sqMeters)
2. Square Miles (sqMiles)
3. Some post processing steps are required. These include updated the final ExcID (EXCID_FNL) and the final Rural field (RURAL_FNL). Also, areas are compared (GIS file versus VP reported sq miles).



b. Input Datasets:

Input Dataset	Description	Source
Exchange Center Centroids	Location of town center for exchanges	Vantage Point, 2010
Rural Exchange Boundaries	Extent of boundaries for rural-based exchanges	Various (Vantage Point, State GIS)
Town Exchange Boundaries	Extent of boundaries for town-based exchanges	Where available, town-based exchanges were extracted from census populated places. Where populated places did not exist, town-based exchanges were digitized from aerial photographs by Stone. 2010

c. *Output Datasets:*

Output Dataset	Description
Rural and Town Exchange Boundaries	Exchange boundaries by state with town and rural extents separated.

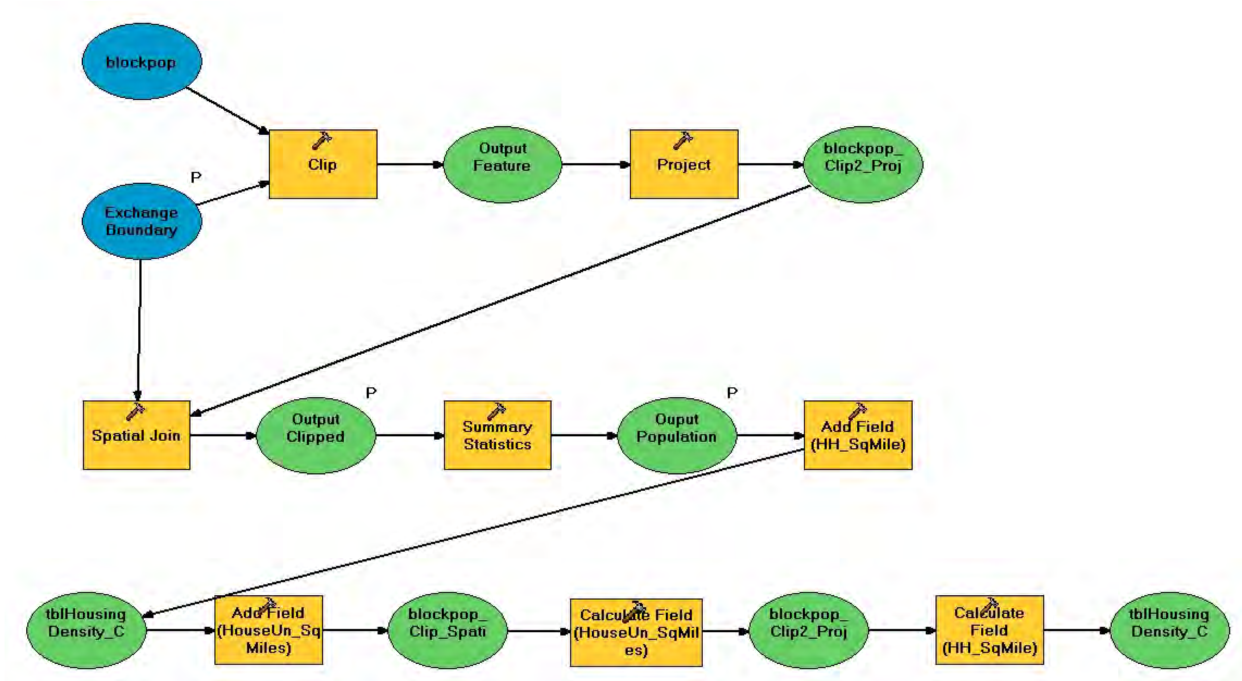
d. *Post-Processing/Clean-up:*

i. Update Field "EXCID_FNL" for rural, town, and all unique ID.
ii. Update Field " Rural_Fnl" for rural, town, and all codes
iii. Delete unnecessary fields
iv. Compare SqMiles (GIS) with VP reported Sq Miles.

2) Housing Density: Tool – 1_HousingDensity

a. *Tool description:* Calculates households and population density per square mile.

i. Extracts 2000 Census block centroids for each exchange
ii. Summarizes total population and households from 2000 Census centroids by exchange
iii. Derives housing density by exchange by calculating households per square mile.



b. *Input Datasets:*

Input Dataset	Description	Source
Rural and Town Exchange Boundaries	Exchange boundaries by state with town and rural extents separated.	Derived from Exchange Boundary Processing tool, various sources.
2000 Census Block Centroids	2000 Census Block Centroid. Contains household, housing unit, and population attributes	ESRI (Census Bureau)

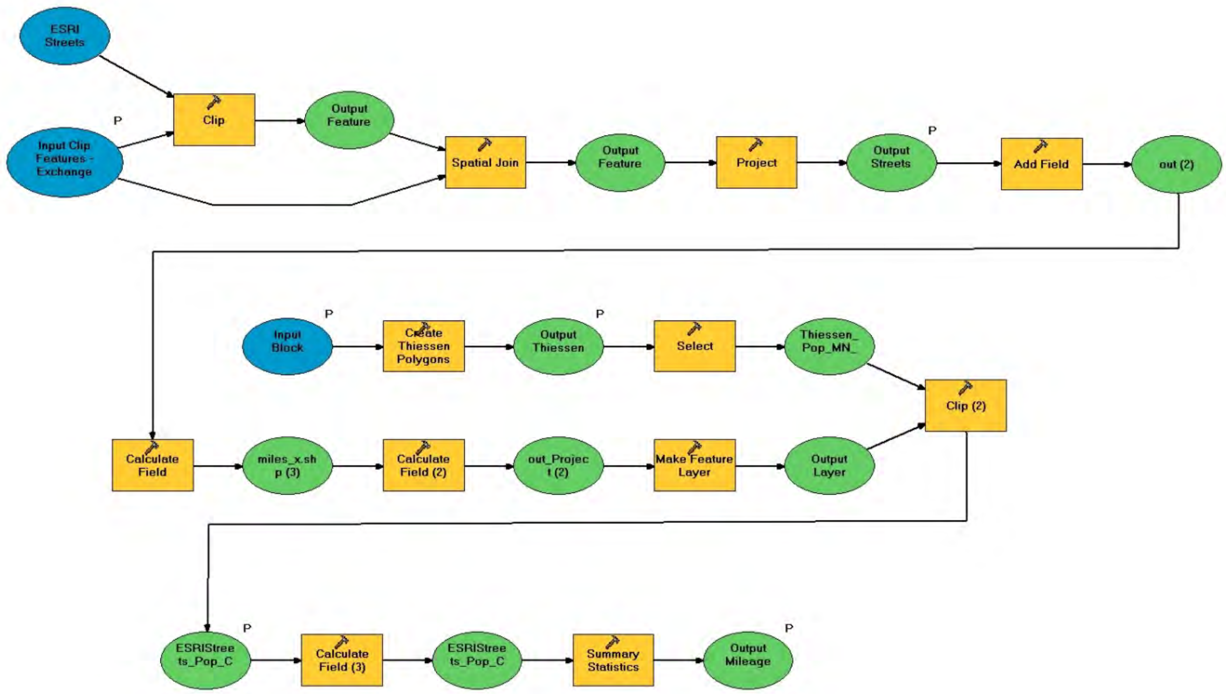
c. *Output Datasets:*

Output Dataset	Description
Output Population Statistics	Table summarizing households and population by exchange
Clipped Block Centroids	Block Centroids clipped to exchange boundaries.

3) Street Mileage: Tool – 2_StreetMileage

a. *Tool description:* Extracts ESRI streets (Streetmap 2010) data for each exchange boundary and calculates mileage by street class. Only streets with population > 0 are included and summarized. Post-processing steps remove roads with FCC road classes 1, 6 and 7

i. Clip ESRI streets to exchange boundary and assign exchange information to street dataset.
ii. Create Theissen polygons from 2000 census block centroids.
iii. Extract ESRI streets that have a population > 0 based on 2000 census block Theissen polygons.
iv. Calculate road mileage by FCC road class for remaining streets with population > 0.
v. Remove FCC classes 1, 6, and 7 from final data layer (Post-processing step)



b. *Input Datasets:*

Input Dataset	Description	Source
ESRI Streemap Streets	Nationwide street dataset (lines). Contains attribute information including FCC road class, road name, among others	2010 ESRI Streetmap (TeleAtlas)
Rural and Town Exchange Boundaries	Exchange boundaries by state with town and rural extents separated.	Derived from Exchange Boundary Processing tool, various sources.
Clipped 2000 Census Block Centroids	2000 Census Block Centroid. Contains household, housing unit, and population attributes	ESRI (Census Bureau), output from Housing Density tool described above

c. *Output Datasets:*

Output Dataset	Description
Clipped ESRI Streetmap Streets	ESRI Streetmap streets clipped to exchange boundaries (all streets).
Census Block Theissen Polygons	Theissen polygons with census block information including population for each exchange

Clipped ESRI Streetmap Streets with > 0 Population	ESRI Streetmap streets clipped to exchange boundaries (>0 population streets).
Output Mileage Table	Summary table of street miles by exchange and FCC road class

d. *Post-Processing/Clean-up:*

i. Clean Clipped ESRI Streets with Population > 0.

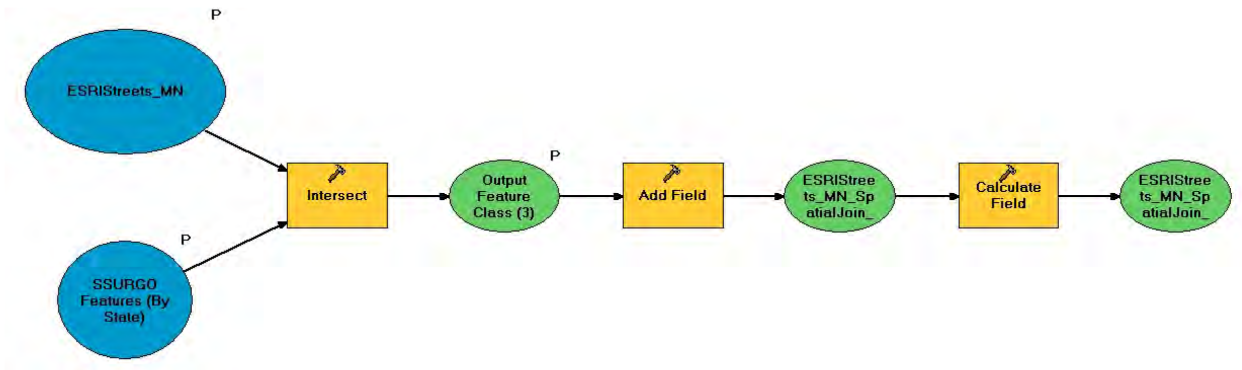
1. Remove streets with FCC classes 1, 6, and 7.
2. Final dataset 'ESRI_Streets_Pop_Filtered'

4) **Prep Soils Layers by State: Tool – 3_Soils_Clip**

a. *Tool description:* Clips SSURGO soils layer to exchange boundaries for each state and intersects with streets for the exchange boundary, resulting in each road segment coded with an MUKEY (Soil code).

i. Intersects ESRI Streets ('ESRI_Streets_Pop_Filtered') with SSURGO soils.

ii. Calculates mileage for each road segment.



b. *Input Datasets:*

Input Dataset	Description	Source
Clipped ESRI Streetmap Streets with > 0 Population	ESRI Streetmap streets clipped to exchange boundaries (>0 population streets).	Derived in 2_StreetMileage tool from step 3 above. 2010 ESRI Streetmap (TeleAtlas)
SSURGO Soil Features	GIS based soil data (polygons) by state	Natural Resources Conservation Service (NRCS) Soil Survey Geographic (SSURGO) database

c. *Output Datasets:*

Output Dataset	Description
ESRI_Streets_Pop_Filtered with Soil MUKEY	ESRI Streetmap streets clipped to exchange boundaries (>0 population streets) with SSURGO soil attribute (MUKEY). FCC road classes 1, 6, or 7 were removed.

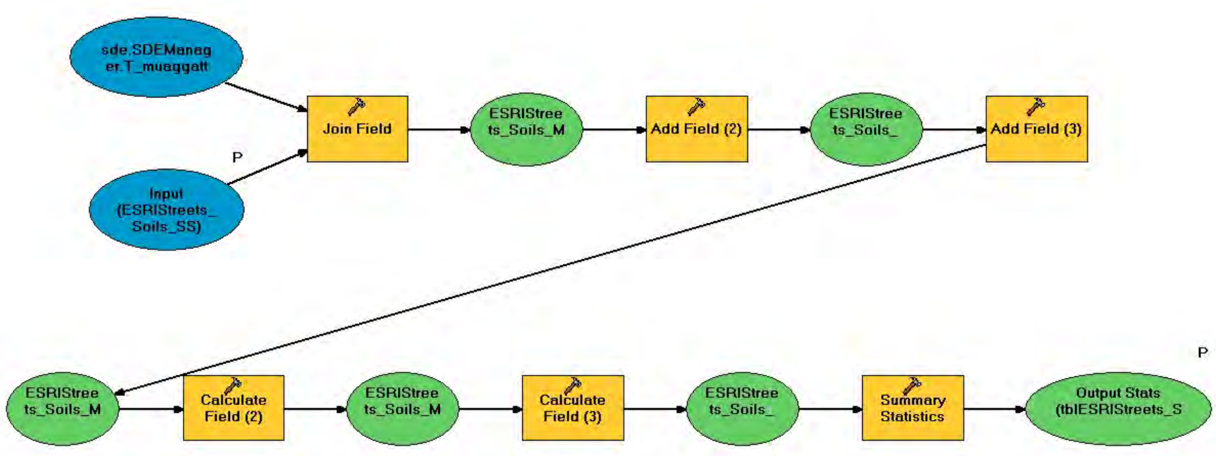
5) Depth to Bedrock from SSURGO Soils: Tool – 4_Soils_BR_GW

a. *Tool description:* Summarizes the miles streets that have soils with a depth to bedrock of < 36".

   i. Joins SSURGO 'muaggatt' table by MUKEY to the 'ESRIStreets_Soils' layer.

   ii. Extracts the attributes for the 'brockdepmin' field from the 'muaggatt' table for each road segment. The 'brockdepmin' field provides 'the distance from the soil surface to the top of a bedrock layer, expressed as a shallowest depth of components whose composition in the map unit is equal to or exceeds 15%' (SSURGO metadata).

   iii. Summarizes the miles of road where SSURGO soils indicate bedrock < 36" from the surface by exchange.



b. *Input Datasets:*

Input Dataset	Description	Source
SSURGO muaggatt table	The muaggatt table is the Mapunit Aggregated Attribute table that records a variety of soil attributes that have been aggregated from the component level to a single value at the map unit level.	Natural Resources Conservation Service (NRCS) SSURGO database
ESRI Streets Pop with Soils	ESRI Streetmap streets clipped to exchange boundaries (>0 population streets) with SSURGO soil attribute (MUKEY). FCC road classes 1, 6, or 7 were removed.	ESRI Streetmap (TeleAtlas) and Natural Resources Conservation Service (NRCS) SSURGO database. Derived in step 4 above.

c. *Output Datasets:*

Output Dataset	Description
tblESRIStreets_Soils_GWBR	Table summarizing the street miles where depth to bedrock is <36”

6) Soils Texture 1: Tool – 5_Soils _RUN_Q1to3_InAccess

a. *Tool description:* Run queries in Microsoft Access that extracts the dominant component and horizon for each soil MUKEY (soil ID) and associated road segment. From the dominant component and horizon of each soil, the representative texture is extracted. From the dominant component, the representative number of frost free days is extracted. The final table provides a texture and the number of frost free days for each MUKEY (SSURGO soil ID).

i. Query 1 – Each soil type and associated MUKEY (Soil ID) have multiple soil components that have various soil properties. For each component, the percent of the component that makes up the soil is provided (‘comppct’ attribute). The component with the highest percent for each soil was extracted in Query 1.

ii. Query 2 – For each component, there are multiple soil horizons in a ‘chorizon’ table. Query 2 extracts the thickest horizon within the top 36” of the soil for the dominant component. Query 2 also extracts the dominant texture for that horizon for each soil and associated MUKEY (Soils ID).

iii. Query 3 – Query 3 extracts the number of ‘frost free days’ from the [ffd_r] attribute for the dominant component of each soil type and associated MUKEY (Soils ID).

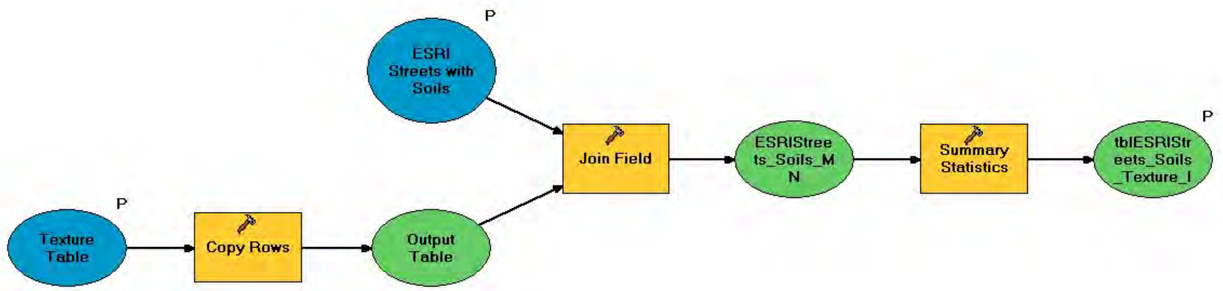
7) Soil Texture from SSURGO Soils: Tool – 6_Soils_Texture

a. *Tool description:* Summarizes the street miles by dominant soil texture class for each exchange.

   i. Joins SSURGO texture table (created in step 6 above) by MUKEY to the ‘ESRIStreets_Soils’ layer.

   ii. Extracts the attributes for the ‘texture’ for each road segment.

   iii. Summarizes the miles of road by dominant soil texture.



b. *Input Datasets:*

Input Dataset	Description	Source
SSURGO texture lookup table	Lookup table with dominant soil texture by MUKEY (SSURGO Soil ID).	Natural Resources Conservation Service (NRCS) SSURGO database. Derived in Step 6 above.
ESRI Streets Pop with Soils	ESRI Streetmap streets clipped to exchange boundaries (>0 population streets) with SSURGO soil attribute (MUKEY). FCC road classes 1, 6, or 7 were removed.	ESRI Streetmap (TeleAtlas) and Natural Resources Conservation Service (NRCS) SSURGO database. Derived in step 4 above.

c. *Output Datasets:*

Output Dataset	Description
tblESRIStreets_Soils_Texture	Table summarizing the street miles by dominant soil texture.

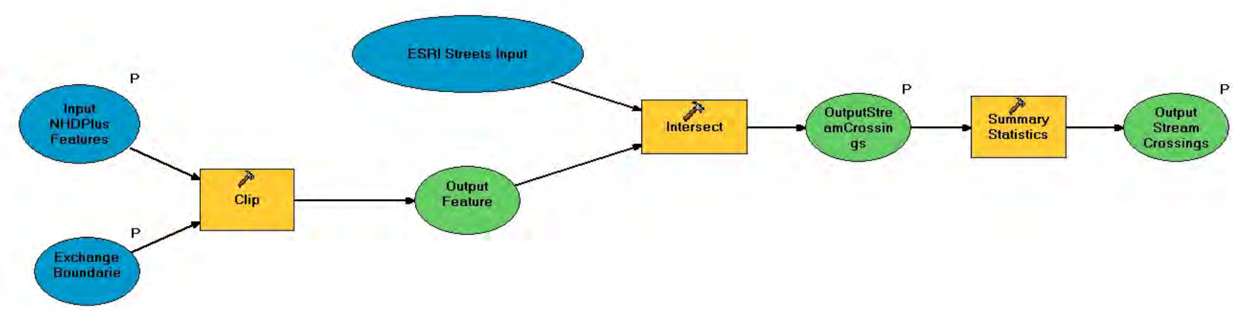
d. *Post-Processing/Clean-up:*
   i. Assigned texture class to each soil texture type based on lookup table provided by Bob Loube. This table was developed by AT&T to characterize soil textures by plow difficulty. The values range from 1 to 3.
   ii. Re-summarized road mileage based on texture class and exchange.

8) Frost Free Days from SSURGO Soils: Tool – 7_Soils_FrostFreeDays

a. *Tool description:* Run query in Microsoft Access that summarize a weighted average frost free days for each exchange.
   i. Query 4 – The representative number of frost free days for each SSURGO soil and associated road segment were extracted in step 6 above. A single value for each telephone exchange was calculated by averaging taking a weighted average of frost free days by road mile.

9) Number of Stream Crossings: Tool – 8_StreamCrossing_NHDPlus

a. *Tool description:* Creates a point where streams intersect with roads and summarizes the total stream intersections by exchange.
   i. Clips stream features to exchange boundaries
   ii. Intersects streams with roads using an output type of 'point'.
   iii. Summarizes the number of stream crossings per exchange.



b. *Input Datasets:*

Input Dataset	Description	Source
Clipped ESRI Streetmap Streets with > 0 Population	ESRI Streetmap streets clipped to exchange boundaries (>0 population streets).	Derived in 2_StreetMileage tool from step 3 above. 2010 ESRI Streetmap (TeleAtlas)
NHDFlowline	The highest resolution national hydrography dataset. NHD Flowline represents stream features from the NHDPlus database	NHDPlus from US EPA

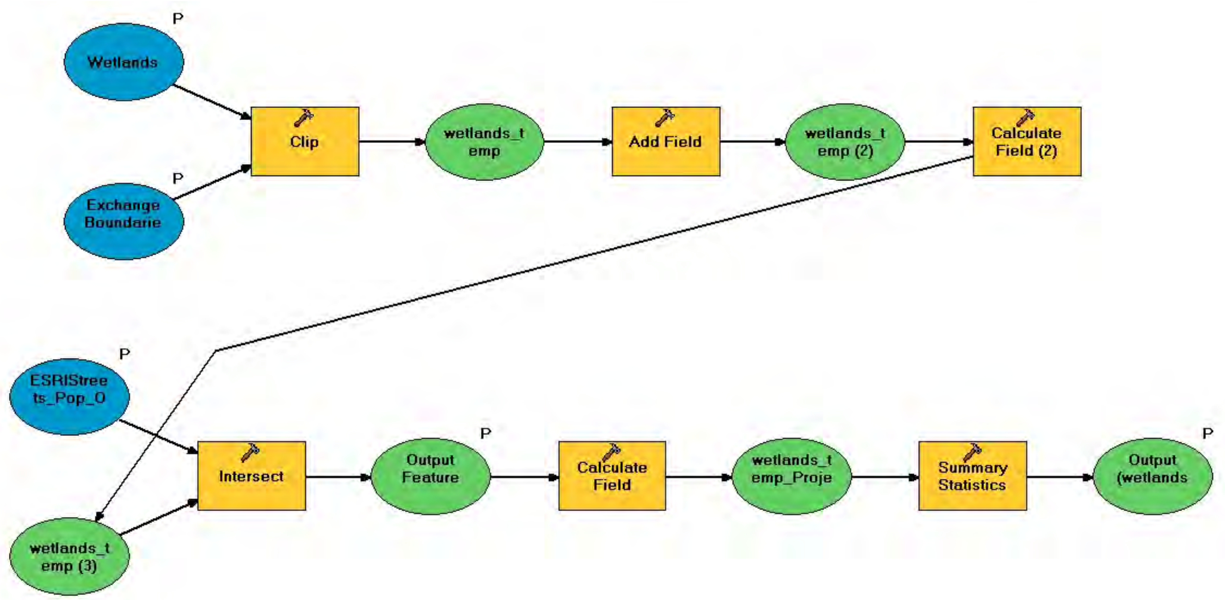
Rural and Town Exchange Boundaries	Exchange boundaries by state with town and rural extents separated.	Derived from Exchange Boundary Processing tool, various sources.

c. *Output Datasets:*

Output Dataset	**Description**
StreamCrossings	Point locations of stream crossings
tblStreamCrossings_Count	Table summarizing the number of stream crossings by exchange

10) Road miles intersecting with wetlands: Tool – 9_Wetlands

a. *Tool description:* Summarizes the street miles that intersect with wetlands.
   i. Clips wetland features to exchange boundaries.
   ii. Intersects the wetlands with ESRI Streetmap streets (Streets with >0 population and excluding FCC road classes 1, 6, and 7)
   iii. Calculates street mileage
   iv. Summarizes the miles of road that intersect wetlands by exchange.



b. *Input Datasets:*

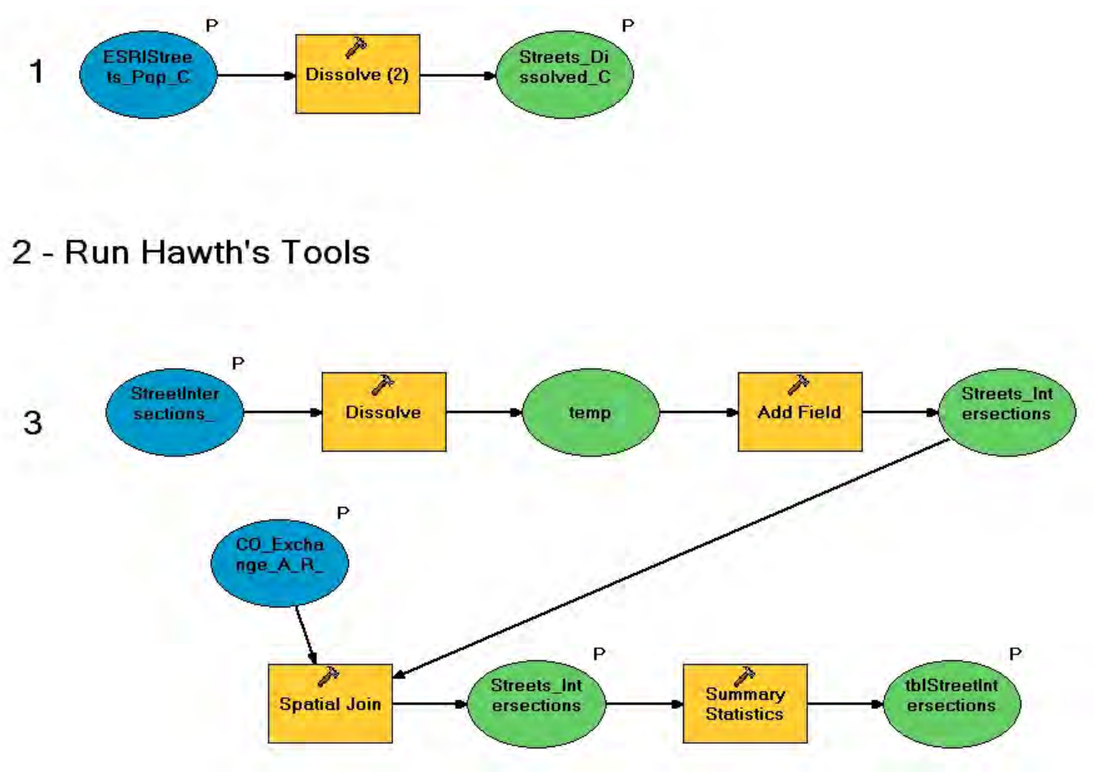
Input Dataset	Description	Source
Wetlands	GIS file with polygons representing wetland areas.	US Fish and Wildlife Service
Clipped ESRI Streetmap Streets with > 0 Population and no FCC classes 1, 6, 7	ESRI Streetmap streets clipped to exchange boundaries (>0 population streets) and no FCC classes 1, 6, 7	Derived in 2_StreetMileage tool from step 3 above. 2010 ESRI Streetmap (TeleAtlas)
Rural and Town Exchange Boundaries	Exchange boundaries by state with town and rural extents separated.	Derived from Exchange Boundary Processing tool, various sources.

c. *Output Datasets:*

Output Dataset	Description
ESRIStreets_Wetlands	Streets that intersect wetlands.
Tbl_Wetlands	Table summarizing the street miles intersecting wetlands for each exchange

11) Number of Street Intersections: Tool – 10_StreetIntersections

a. *Tool description:* Creates a point where streets intersect and summarizes the total number of street intersections by exchange.
   i. Dissolves Streets by Street Name
   ii. Creates points for street intersections by using the 'Intersect Lines' tool within the 'Hawth's Tools' toolbar.
   iii. Assigns exchange information to street intersection points.
   iv. Summarizes the number of street intersection by exchange.



b. Input Datasets:

Input Dataset	Description	Source
Clipped ESRI Streetmap Streets with > 0 Population and no FCC classes 1, 6, 7	ESRI Streetmap streets clipped to exchange boundaries (>0 population streets) and no FCC classes 1, 6, 7	Derived in 2_StreetMileage tool from step 3 above. 2010 ESRI Streetmap (TeleAtlas)
Rural and Town Exchange Boundaries	Exchange boundaries by state with town and rural extents separated.	Derived from Exchange Boundary Processing tool, various sources.

c. Output Datasets:

Output Dataset	Description
Dissolved ESRI Streets	ESRI Streetmap streets clipped to exchange boundaries (>0 population streets) and no FCC classes 1, 6, 7 and dissolved by street name
StreetIntersections	Point locations of street intersections
TblStreetIntersections	Table summarizing the number of street intersections by exchange

12) **Precipitation: Tool – 11-Precip_Greater_0.5Inches**

a. *Tool description:* Use precipitation data from the National Climate Data Center to extract the average number of days with > 0.5 inches of rain for each exchange using Microsoft Access queries.

  i. Precipitation data was purchased from NOAA National Climate Data Center (NCDC)

  ii. The nearest SAMSON station was identified for each exchange boundary using SAMSON station location data from the GeoStac database.

    1. GeoStac: http://geostac.tamu.edu/
    2. SAMSON is a compilation of National Renewable Energy Laboratory solar data and National Weather Service hourly surface observations.

  iii. The precipitation data was extracted for each SAMSON station (associated COOP station).

  iv. Precipitation data was imported into an Microsoft Access database:

    3. K:\USA\Weather\NOAA\FiberToHome_Precip\NCDC_SummofDay_Precip.accdb

  v. Precipitation data was transposed using a Microsoft Access module.

  vi. Precipitation data was summarized and averaged over the 25 year period from 1975 to 2000 to extract number of days with > than 0.5 inches of precipitation per year. Data for these years was available for all stations.

    4. Microsoft Access Query: qryNumberOfDaysGrtrThan5in_2

Signed: Katie Budreski

Date: 12/23/2010